

Key characteristics of specular stereo

Alexander A. Murry^{1,2}, Roland W. Fleming³ & Andrew E. Welchman^{1,4}

¹ School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

² Department of Psychology, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK

³ Department of Psychology, University of Gießen, Otto-Behaghel-Str 10/F, 35394, Gießen, Germany

⁴ Department of Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, UK; correspondence: aew69@cam.ac.uk

Accepted at *Journal of Vision* 21st October 2014

Because specular reflection is view dependent, shiny surfaces behave radically differently from matte, textured surfaces when viewed with two eyes. As a result, specular reflections pose substantial problems for binocular stereopsis. Here we use a combination of computer graphics and geometrical analysis to characterize the key respects in which specular stereo differs from standard stereo, to identify how and why the human visual system fails to reconstruct depths correctly from specular reflections. We describe rendering of stereoscopic images of specular surfaces in which the disparity information can be varied parametrically, and independently of monocular appearance. Using the generated surfaces and images, we explain how stereo correspondence can be established with known and unknown surface geometry. We show that even with known geometry, stereo matching for specular surfaces is non-trivial because points in one eye may have zero, one or multiple matches in the other eye. Matching features typically yield skew (non-intersecting) rays, leading to substantial ortho-epipolar components to the disparities, which makes deriving depth values from matches non-trivial. We suggest that the human visual system may base its depth estimates solely on the epipolar components of disparities while treating the ortho-epipolar components as a measure of the underlying reliability of the disparity signals. Reconstructing virtual surfaces according to these principles reveals that they are piece-wise smooth with very large discontinuities close to inflection points on the physical surface. Together, these distinctive characteristics lead to cues that the visual system could use to diagnose specular reflections from binocular information.

Keywords: Stereopsis, Matching, Correspondence Problem, Binocular Vision, Specularity, Material Perception.

1. Introduction and goals

Binocular vision provides humans and machines with a ready source of information about the depth structure of a surrounding scene. To infer depth from binocular disparities, it is first necessary to match image features between the two views. For matte objects, elements that match between the viewpoints tend to be similar in form, arise at similar locations in the image (at least vertically), and vary smoothly across space. However, specular objects (such as a polished kettle or chrome bumper) can give rise to binocular disparity signals quite different from those that arise from matte objects. Here, we aim to explain and detail these differences.

A well-known feature of specular reflections is that they often lie at a location in space that is displaced from the true surface of the object (Blake & Bülthoff, 1991; Hurlbert, Cumming, & Parker, 1991; Kerrigan & Adams, 2013). This contrasts with non-specular objects where disparity values map to surface depth in a straightforward way. This difference in the relationship between depth values and surface location poses a potential challenge to both artificial and human visual systems. Here, we seek to characterize the conditions that lead to the displacement of specular reflections. We do this as a means of understanding why human observers treat reflections as though they are true surface markings when judging depths (Murry, Welchman, Blake, & Fleming, 2013). In the process, we also quantify other respects in which specular reflections deviate from standard stereopsis.

Most previous theoretical and computational work has focussed on the behaviour of individual highlights (Koenderink & van Doorn, 1980; Longuet-Higgins, 1960) or surface reconstruction from multiple images (including stereopsis and movement): (Blake & Brelstaff, 1988; Oren & Nayar, 1997; Sankaranarayanan, Veeraraghavan, Tuzel, & Agrawal, 2010; Vasilyev, Adato, Zickler, & Ben-Shahar, 2008; Vasilyev, Zickler, Gortler, & Ben-Shahar, 2011; Zisserman, Giblin, & Blake, 1989). However, few of these studies explicitly spell out the main challenges that specular stereo present to the human

visual system. Here we characterize in detail several key properties of specular stereopsis. First, we present a method for determining ground-truth stereo matches for mirror surfaces of *known* geometry, demonstrating the presence of image regions for which meaningful stereo matches do not exist. Then, we describe key features of specular disparities that are potentially important for both biological and machine stereo-vision. In particular, we detail the presence of non-epipolar disparity matches and the potential for very large disparity gradients and discontinuities. We further address the instability of specular disparity fields with respect to variations of viewing/surface geometries. Finally, we show that the distribution of ortho-epipolar disparities is related to surface geometry, providing a constraint when estimating the curvature of the viewed object. Thereby we show that even though specular stereo signals do not support direct perceptual estimates of the physical shape of an object (Murry, Welchman, Blake, & Fleming, 2013), specular disparity fields do carry information about the intimate relations between the viewing geometry and surface topography which could potentially be exploited by humans and artificial systems.

2. Specular and Lambertian illumination mapping

To frame the problem of specular stereo and its differences from the typical case of a Lambertian object, we start by considering the ray geometry of binocular image generation. We pose this as the process of generating a computer image; however, the exposition describes the information that is available to either a human or artificial visual system. Understanding the relevant image information is a crucial step for determining which binocular cues the human visual system could use to identify and interpret the disparities produced by specular surfaces.

Rendering an image of an object depends on three main elements: (i) the object's surface (geometry and material), (ii) the viewing geometry (left and right viewpoints and orientations) and (iii) the illumination provided by the scene

(from a single point light source to a full illumination map). Let us assume that (i) the surface S and its normal vectors N are known, and the surface material is a perfect mirror¹, (ii) the viewpoints are located at E_L and E_R (at finite distance from the surface) and (iii) the illumination is a spherical illumination map at infinity (Debevec, 2008). To render the image of the object entails that for each visible point P of surface S we determine its pixel value in the images for E_L and E_R .

2.1. The rendering process for an ideal mirror

Ideal mirrors do not have texture markings, so the image consists of nothing more than a distorted reflection of the surrounding environment. In order to find the pixel value of point P in the image of eye E_L , we trace viewing vector $\mathbf{v}_L = P - E_L$, calculate the reflected ray vector by the law of specular reflection $\omega_L = 2(\mathbf{n} \cdot \mathbf{v}_L)\mathbf{n} + \mathbf{v}_L$, trace it out to the environment and take the corresponding pixel value of the spherical illumination map (Fig 1b).

There are three main observations to make about this generative process. First, notice that since viewing vectors from left and right eyes at surface point P are different $\mathbf{v}_L(P) \neq \mathbf{v}_R(P)$, the left and right reflected ray vectors cannot be equal $\omega_L(P) \neq \omega_R(P)$ such that different locations in the environment are viewed by the two eyes. Therefore, surface point P will generally have different pixel values in E_L and E_R eyes and thus left and right images of this surface point will be disparate. Second, notice that the reflected ray vectors that determine the pixel values vary twice as fast as surface normal vectors. This leads to the characteristic distortions of the environment in mirror reflections (under orthographic projection a hemispherical mirror images the entire sphere around the surface).

Third, near² surface inflection points where the surface normal vectors start turning in the opposite direction, the reflected vectors invert. In consequence the reflected ray vectors sweep through the same portions of the environment several times, giving rise to multiple reflections of the same feature in the environment map (see Fig 1b stereo pair: the reflection of Utrecht Dom Tower appears three times). This multiple mapping of the environment to the image has the potential to give rise to significant confusion when calculating stereo correspondence. Nevertheless, in Section 3 (below), we describe how local matching could in principle be used to filter out such potentially misleading matches to ensure that matches come from surface patches with qualitatively similar structure. In particular, although global matches (e.g. three images of the Dom Tower) involve reflections of the same portions of the surrounding environment, from the point of view of stereopsis, two should be considered spurious matches (like those that occur in the wallpaper illusion) and thus be filtered out.

2.2. Rendering surfaces with specular reflections stereoscopically 'painted' onto the surface

Being able to render a graphical image of a specular object is a good starting point to understand the way in which the human visual system might process the information it contains. However, from an experimental perspective it is useful to be able to construct versions of the stimuli that differ along a key dimension (e.g. specularity), while keeping other factors (e.g., low level image statistics) as similar as

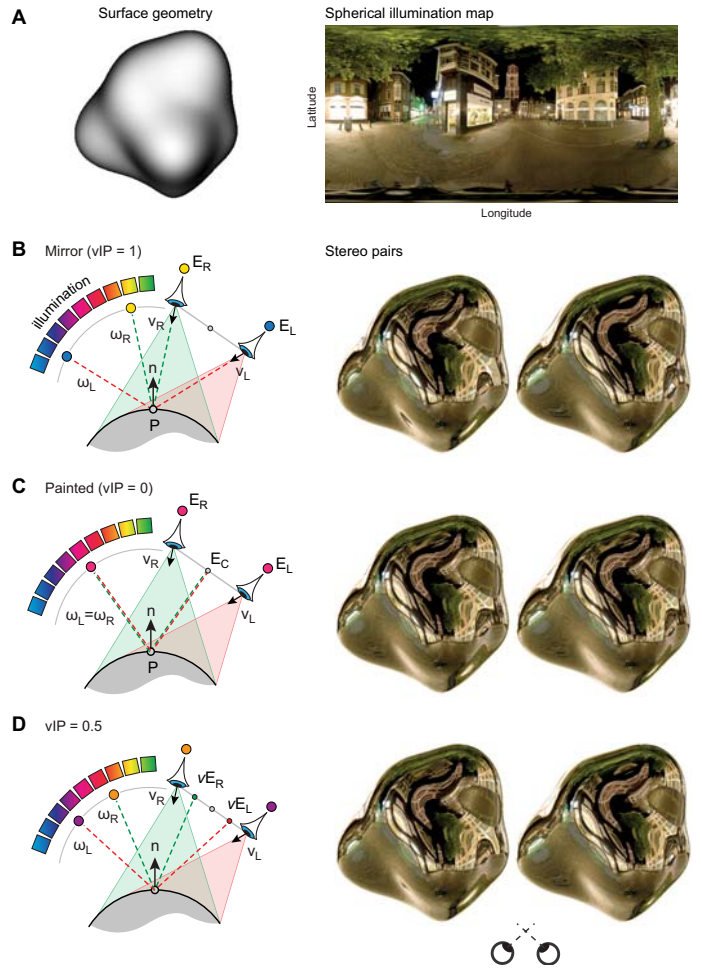


Figure 1. The stereo rendering process. A) Creating stereo images of reflective objects involves a 3D shape model (left) illuminated by a spherical illumination map (right). Here the illumination map is unwrapped into a latitude-longitude projection. B) The rendering process for a mirror. Point P on the surface of the object is viewed from eyes E_R and E_L . The pixel value at point P is determined by the reflection of the view vectors ($\mathbf{v}_R, \mathbf{v}_L$) around the surface normal (\mathbf{n}) at point P . The reflected ray vectors ω_L and ω_R point to different locations in the illumination map, meaning that location P has different pixel values in the two images. This is shown schematically by the rainbow illumination map and the dots behind each eye. Stereograms (right) are presented for cross-fusion. C) The rendering process for a 'painted' shape (virtual illumination point, vIP = 0). Here the pattern of reflections is determined using a view ray from the cyclopean point (E_C). Tracing out rays from E_C across the whole surface produces characteristic specular distortions, which are then imaged binocularly from the two viewpoints. Note that the stereoscopic frustum is the same as in B, the only difference is the location from which pixel intensities are determined. D) Manipulating the virtual illumination point. Pixel intensities can be determined from any location along the interocular axis. Here the points from which to determine reflections are half way between the eye positions and the cyclopean point.

possible. Previous studies on the role of individual point light highlights have used such an approach by placing single highlights on the surface of the object (Blake & Buelthoff, 1990; Wendt, Faul & Mausfeld, 2008). Here we use a similar approach for full scene reflections and thereby isolate and characterize the effects of specularity on human stereopsis (see section 6). In particular, we can quantify the differences

¹ Our analysis and observations are based on rendering an ideal mirror (excluding inter-reflection). While this is a simplification, the specular and diffuse components of a partially specular surface can to a first approximation be treated independently, yielding two distinct disparity fields.

² It is near, rather than exactly at, surface inflection points because the image creation process depends on the combination of surface and viewing geometry

between specular and matte versions of an object to understand the specific properties of specular stereo available to the human visual system, and then use such stimuli to conduct perceptual experiments. Here we describe a way in which such stimuli can be generated in order to isolate the binocular differences between matte and specular objects while the monocular content of the stereo-pairs remains as close as possible to identical across mirrored and matte versions. This is important for our subsequent analysis as it allows us to directly compare the binocular properties of matte and specular surfaces.

As a starting point, we can ‘paint’ the specular reflections onto the surface of an object so that the reflections are practically indistinguishable from true mirror reflections when viewed monocularly, but which have all the disparity characteristics of standard surface texture markings when viewed binocularly. Such ‘painted’ stimuli are akin to ‘sticky’ reflections for moving objects (Doerschner et al., 2011). Here we show how the ‘painting’ approach can be generalised to allow parametric manipulation of specular objects suitable for studies of the human visual system. This complements suggestions for perceptually-motivated “short-cuts” when rendering specular- (Templin, Didyk, Ritschel, Myszkowski, & Seidel, 2012) and refractive- (Dąbala et al., 2014) objects.

In order to make left and right images stereoscopically consistent, the pixel values of the surface should be independent of viewing point, i.e. the reflected patterns should be attached (‘painted’ on) to the surface. We also would like to be able to ensure that monocular properties of the images are similar to mirrors, i.e. they should have mirror-like distortions. To achieve this, we map the environment onto the surface using reflected ray vectors cast from the cyclopean point $E_C = (E_L + E_R) / 2$ (Fig. 1c). Notice that there is no camera (i.e., no image formation) at the cyclopean point. Rather, it is used only for mapping (‘painting’) the environment onto the surface when images of the surface are rendered from eyes E_L and E_R , (i.e. the view vectors used for rendering the stereopair do not change). Since the mapping process is governed by the laws of specular reflection, the images will have distortions similar to mirrors. However, the mapping does not depend on the true position of the eyes and therefore each surface point will have the same pixel values in left and right images and can thus be matched stereoscopically. Moreover, as we describe next, this approach can be generalized to create stereopairs whose disparity properties vary continuously between mirror-like and standard surfaces, while keeping the monocular properties of the image almost constant.

2.3. Virtual illumination mapping

Using the logic of the rendering approach described above, we can construct artificial stereo-images whose disparity properties range smoothly between mirror and matte/textured surfaces. For the left eye we map illumination onto the surface using point vE_L (virtual illumination point for left eye) and vE_R for right eye (Fig. 1d). Virtual illumination mapping points are placed on inter-ocular axis equidistant from cyclopean point

$$vE_L = E_C + vIP * (E_L - E_C) / \|E_L - E_C\|$$

$$vE_R = E_C + vIP * (E_R - E_C) / \|E_R - E_C\|$$

where vIP is virtual illumination point index. Note again that the virtual illumination points are used only for mapping environment on the surface while actual images are taken from real viewpoints E_L and E_R , i.e. viewing vectors are fixed. Condition $vIP = 0$ corresponds to ‘painted’ case described above (virtual illumination points coincide at the cyclopean

point). If $vIP = 1$ then virtual illumination points coincide with the true locations of the corresponding viewpoints, leading to standard mirror reflections. Varying vIP smoothly allows us to construct stimuli with stereo-properties ranging between (or indeed beyond) these two extremes.

This technique enables stimuli to be generated from a parametric space of disparity-defined objects. Figure 1 provides examples of three stimuli drawn from the vIP space, while in Figure 2 we quantify how the displacement of the highlights from the physical surface varies as a function of this manipulation to show how the range of depths in the objects changes as vIP is manipulated. In this paper, we focus on using such stimuli to quantify the type of stereoscopic information available to viewers. However, empirically, the vIP space lends itself to systematic testing of human judgments of shape and material. In particular, the ability to systematically vary binocular signals while keeping monocular information more-or-less constant could be exploited to understand the weighting process by which monocular and binocular information is combined when observers make judgments about 3D shape and material properties. Here, we rely on this manipulation to characterize the key differences between specular and non-specular disparity signals.

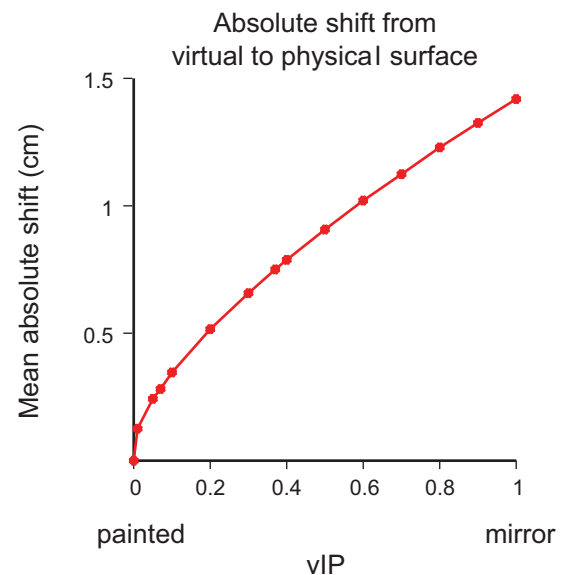


Figure 2. Quantifying the effect of manipulating the virtual illumination point on the divergence between the physical surface and the virtual surface described by binocular specular reflections. The graph shows the mean unsigned depth offset between the physical and virtual surfaces for four ‘potato’ objects (spheres randomly perturbed by 100 Gaussian blobs) as vIP was manipulated. Viewing distance was 50 cm, interocular separation 6.5 cm, and the objects were approximately 7 cm in diameter – i.e., like looking at an apple or potato at arm’s length. Depth displacements greater than 10 cm were only found to originate from unfusable image locations; we therefore treated them as outliers in calculating the mean offset value. The vIP manipulation causes a systematic, regular and monotonic change in the depths of the stimulus.

It should be noted that (Dąbala et al., 2014) recently presented a similar approach to manipulating stereoscopic signals for reflective objects. Their approach was designed to promote “visual comfort” in the displays, using manipulations equivalent to the vIP for each rendered pixel in the image (meaning that individual pixels are rendered with from different virtual illumination points).

3. Determining specular stereo-matches for an object of known shape using ray geometry

In Section 2 we considered the forward process by which the environment is mapped to left and right eye images. Here we describe the disparity field that results when viewing specular objects binocularly, by solving the correspondence problem for the simpler case of known geometry. Although of course, stereopsis usually deals with *unknown* geometry, in order to frame the problem correctly, and to establish ground-truth estimates of specular disparity fields, it is useful first to consider the case with known geometry. In particular, these ground truth descriptors are based on a forward model of disparity generation that exploits the known geometry of the viewed shape. This provides a purely geometric definition of the available disparity information that is independent of the content of the illumination field. As we shall see, and unlike the case of standard textured surfaces, for specular surfaces even when geometry is known, establishing correspondence is not entirely trivial.

To calculate the disparity field we need to solve the stereo correspondence problem: for each location in the left eye's image seek the location in the right eye's image that has the same generative cause. For a matte, textured object, we can think of this process as two eyes viewing a particular texture element of the surface of the object, with the brain charged with establishing correspondence between the retinal projections of the texture elements in the two images. In the specular case, we assume (like previous work on specular stereo, (Murry et al., 2013)) that the definition of correspondence is essentially equivalent; namely, that the visual system seeks the image feature in one eye that matches the same image feature in the other eye. Thus, the basic task is the same (seeking correspondence between image features), however, in the specular case, these features originate from the reflection of the environment illumination map, rather than markings on surface itself. Given this definition, corresponding points in the environment yield the same pixel values in the two images (up to sampling limits).

It is important to note that defining correspondence in terms of matching image features means that, in general, the resulting disparities do not lie on the surface in depth (we explain this observation in detail in Section 3.1). To correctly reconstruct surface depths would require finding the projections of corresponding *surface* features in the two eyes' views (e.g., matching extrema of curvature). However, the definition of correspondence in terms of matching image features makes more sense given the optics of mirror reflection. Reflections are *virtual images*, whose location in 3-space is specified by the geometry of reflection. Importantly, the depth of the virtual images is therefore consistent across all parallax-based depth cues (stereopsis, motion parallax, accommodation). For example, in order to bring the reflections into focus, it is necessary to focus not at the distance of the surface but at the location behind or in front of the surface that is consistent with the disparity signals³. Matching corresponding *surface* features would require favouring matches that have large inter-ocular differences in the image values, while suppressing the much better image matches from the virtual image. Given that this is the exact opposite of normal stereopsis—for which the visual system is presumably optimized—it seems intuitively unlikely that the visual system would prefer surface matches to image matches, even though it is the surface matches that would indicate the true physical location of the surface. This intuition is supported

by our previous findings that when subjects are asked to report the perceived depths of surface locations, they generally report depths that are much closer to the virtual surface than to the true physical surface (Murry et al., 2013), implying that the visual system does indeed match image features rather than surface locations.

While the principles of establishing correspondence are straightforward, pixel intensity *per se* is not a generally useful characteristic for matches because (i) the environment map may contain repetitive pixel values (e.g. a picket fence) that correspond to unrelated reflected ray directions and (ii) surface concavities entail that multiple surface locations can reflect the same portion of the environment map (e.g. the Dom in Fig. 1). To define a unique match for each image location we therefore need to constrain the solution, the logic for which we now describe.

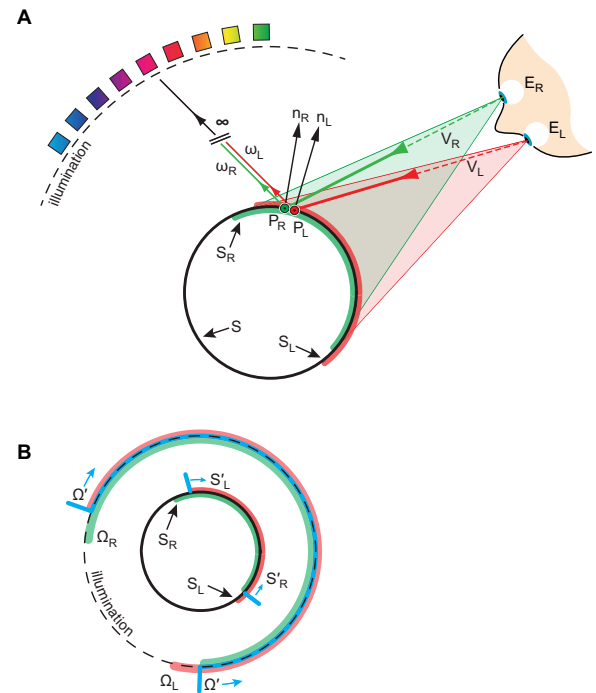


Figure 3. Establishing stereo correspondence. A) Calculating binocular disparities depends on matching locations that point to the same place in the illumination map. Here, points P_L and P_R of surface S reflect same portion of the environment to eyes E_L and E_R . This correspondence can be identified by finding reflected ray vectors ω_L and ω_R that are parallel (note that this occurs even though the normals n_L and n_R are different, because of the difference of view position). Notice that different portions of the surface (S_L , S_R) are visible to the two eyes – denoted by the shaded regions around the surface. B) The differences in the visible portions of the surface mean that different portions of the illumination map are visible to the two eyes, leading to unmatchable features. This is described as the set of reflected ray vectors Ω_L , Ω_R . The intersection of these reflected ray vectors (Ω') defines the space within which binocular correspondence can be established.

3.1. Reflections depend on the viewpoint, thus the same surface point P reflects different portions of the environment to left and right eyes.

To start we deal with the simplified one-dimensional case (i.e. a cross section through a shape; Fig. 3), then we expand to finding correspondence in two dimensions. Given a surface of known shape S with known normal vectors $\mathbf{n} = \mathbf{n}(P)$, $P \in S$, we can define S_L and S_R as the portions of the surface that are visible to left and right eyes E_L and E_R . Con-

³ Recall that for ideal mirrors, reflections are the *only* visible features.

sider point P on the surface S , the images of point P in the left and right eyes are defined by the reflected ray vectors ω_L and ω_R (see **Fig. 1**). According to the law of specular reflection:

$$\omega_L = 2 (\mathbf{n} \cdot \mathbf{v}_L) \mathbf{n} + \mathbf{v}_L$$

$$\omega_R = 2 (\mathbf{n} \cdot \mathbf{v}_R) \mathbf{n} + \mathbf{v}_R$$

where $\mathbf{n} = \mathbf{n}(P)$ is the unit surface normal vector ($\|\mathbf{n}\| = 1$) at point P , and \mathbf{v}_L and \mathbf{v}_R are normalized left/right viewing vectors:

$$\mathbf{v}_L = (P - E_L) / \|P - E_L\|$$

$$\mathbf{v}_R = (P - E_R) / \|P - E_R\|.$$

Since the eyes are separated $E_L \neq E_R$, for a real mirror surface, the left and right reflected vectors cannot coincide $\omega_L(P) \neq \omega_R(P)$, i.e. left and right reflected vectors must point at different locations in the environment. Thus, every surface point P forms different images in left and right eyes. To solve the stereo-matching problem we need to find such points P_L and P_R whose corresponding reflected ray vectors point to the same location in the environment and thus form the same pixel values in left and right images (**Fig. 3a**). Assuming the environment is infinitely far away it is sufficient to identify reflected ray vectors that are parallel. In other words, for each point $P_L \in S_L$ (portion of S that is visible to E_L) we need to find a point $P_R \in S_R$ such that $\omega_L(P_L) = \omega_R(P_R)$ (or $\omega_L \times \omega_R = 0$). Treating the environment as infinitely far is a reasonable simplification; our analysis (see Appendix 1) shows that under normal conditions this assumption should not lead to significant depth errors. For some portions of the surface, solutions do not exist while for non-convex shapes there may be multiple solutions. We address such situations below.

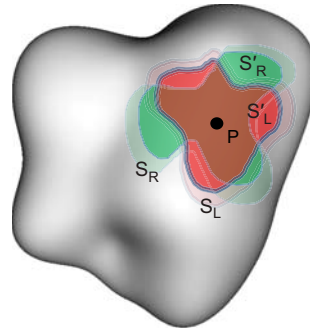
Rather than considering a single reflected ray, let us move on to consider the set of all possible reflected ray vectors. In principle, the space of reflected ray vectors may cover the entire sphere of possible directions (if the viewpoint is infinitely far from the object). However generally the reflected ray vectors will occupy only a subspace of the sphere of possible directions. The subspaces of reflected vectors for left $\Omega_L = \Omega(S_L)$ and right $\Omega_R = \Omega(S_R)$ eyes do not completely coincide, i.e. $\Omega_L \neq \Omega_R$ (however there is considerable overlap, **Fig. 3b**). It is clear that the stereo-matching solutions exist only for points whose reflected ray vectors overlap. In other words, the solution to the stereo correspondence problem exists only for surface points from $S_L' \in S_L$ and $S_R' \in S_R$ where S_L' and S_R' are such that $E_L:S_L' \rightarrow \Omega'$ and $E_R:S_R' \rightarrow \Omega'$ where $\Omega' = \Omega_L \cap \Omega_R$. Each point of S_L' must have a corresponding stereo-match S_R' , i.e. $\forall P_L \in S_L' \exists P_R \in S_R' : \omega_L(P_L) = \omega_R(P_R)$ where $\omega_L \in \Omega'$, $\omega_R \in \Omega'$. Thus we have a formulation for the regions within which to establish stereo correspondence.

The corollary of this is that for points outside S_L' and S_R' stereo-matches do not exist. In the simple case of a sphere, this absence of correspondence is similar to 'da Vinci occlusion' (Nakayama & Shimojo, 1990) where the edges of a solid objects are differentially visible for the two eyes. Note however that there is an important difference for the specular case in that these areas are more pronounced because reflected vectors vary faster than surface normals. However, as we discuss next, for surfaces that have concave regions, areas of missing stereo correspondence are not limited to the physical edges of the object.

Having illustrated the problem in a one-dimensional slice, we now move to the two-dimensional case (**Fig 4**). If a viewed reflective surface has concavities, the reflected ray vectors are not unique because different surface points can reflect the same portion of the environment (e.g. the Dom in **Fig 1**). In consequence, the global solution will generally not be unique, i.e. for single P_L there may be multiple stereo-matches $P_{R1}, P_{R2}, \dots, P_{Rn}$ such that $\omega_L(P_L) = \omega_R(P_{R1}) = \dots = \omega_R(P_{Rn})$; this poses a challenge in deciding which match to choose. We suggest taking a *local* match whereby corresponding points P_L and P_R belong to a smooth, surface patch.

Let us consider point P on the surface and construct around it patches S_L and S_R , which project uniquely to the space of reflected ray vectors Ω_L and Ω_R (**Figure 4**). The edges of these patches will be very close to (although not coincident with – see footnote 2) the inflection contours of the surface where the sign of curvature changes and thus the normal vectors reverse. Beyond the boundaries there are areas where no stereo solution exists because of 'da Vinci' like differences in the portions of the environment visible to the two eyes; note however that these regions are not limited to a shape's physical boundaries but can occur in the centre of the visible portion of the shape. In other words, inflection contours naturally divide a non-convex specular object into

Projection onto surface



Space of reflected ray vectors

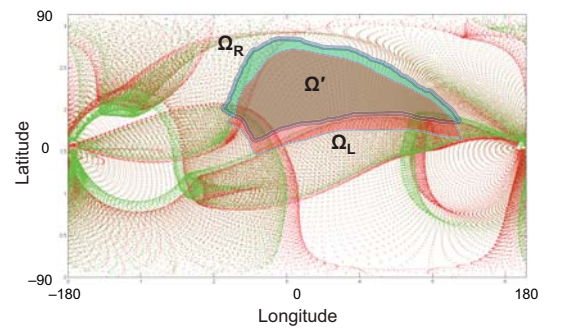


Figure 4. Finding correspondence in two dimensions. We can construct surface regions around point P for which stereo solutions exist. Portions S_L and S_R of surface S are visible to eyes E_L and E_R , and they reflect portions Ω_L and Ω_R of environment Ω . Their intersection of $\Omega' = \Omega_L \cap \Omega_R$ contains reflected ray vectors that are visible to both eyes, thus defining the space within which to identify stereo matches. Defining this surface patch provides a local region within which to identify correspondence: for each point of S_L' there must exist a specular stereo-match in S_R' , where S_L' and S_R' are portions of surface S which reflect Ω' to E_L and E_R .

patches with locally smooth disparity fields, which are separated by regions of unmatchable features, for which disparity is undefined. Such regions create difficulties for machine and human vision as stereo cues specify internal contours where no depth is defined despite monocular image features being contiguous. We illustrate the presence of these internal boundaries in the disparity field in **Fig 5**, where we constructed 'painted' and 'specular' stereograms of a 3D shape reflecting an illumination of uniformly sized spheres. This illumination map provides a clear illustration of the way in which environmental features are distorted by reflections – changing the isotropic illumination into patches with local orientations. For the specular case, the different environmental features visible to the two eyes produce locations where binocular disparity is undefined. These internal contours divide the shape into a series of smooth islands where disparity is defined.

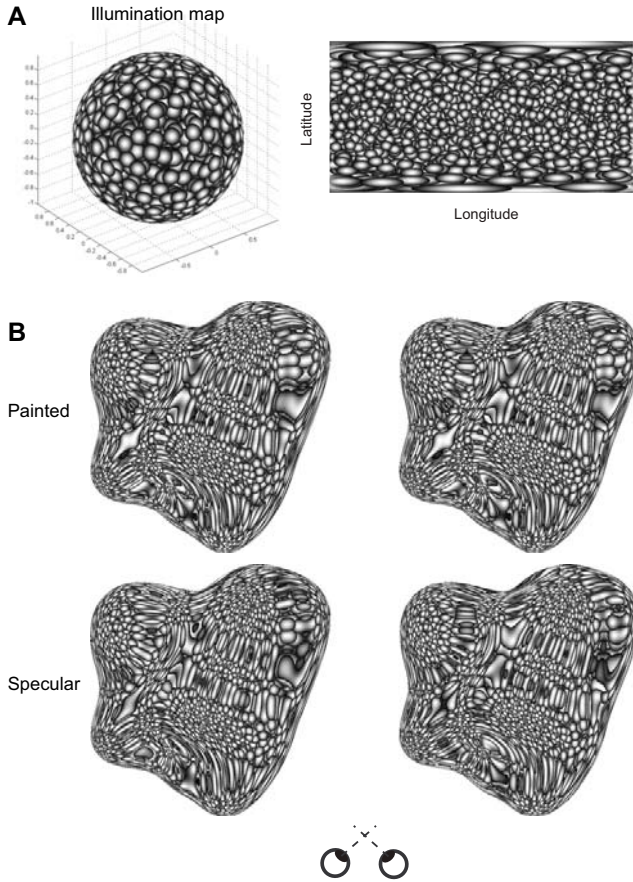


Figure 5. Illustration of piece-wise smoothness of the disparity field. We rendered a 3D object with concavities under an isotropic illumination map containing spheres. This allows a clear visualization of the distortions introduced by specular reflection – i.e. regions in which there is a rapid change in the reflection vectors result in elongated features on the surface of the object. These regions align with piece-wise smooth patches for an object with a specular surface. Outside these ‘islands’, disparities can become very large and are often undefined. Stereograms are presented for cross-fusion.

While the surface patches S_L and S_R overlap considerably, they do not coincide perfectly because the reflected vectors depend not only on surface normals but also on the viewing vectors, which are different for left and right view-points. In consequence, we can find the overlap of reflected ray vectors $\Omega' = \Omega_L \cap \Omega_R$ and then project this overlapping region back into left/right surface patches S_L' and S_R' (Fig. 4). For each point in S_L' there must exist a unique stereo-match in S_R' and through surface smoothness, this mapping must be continuous and the corresponding disparity field should also be smooth. Thus, conceptualising the reflected ray vectors in this manner allows us to ensure a local match where corresponding points arise from surface regions with similar topological properties between the two eyes. While other global matches are possible (e.g. other copies of the Dom Tower), such matches would generally belong to a surface patch with qualitatively different surface structure and thus such matches would not provide useful information about local surface geometry. Moreover, such matches would cross inflection contours, often resulting in large binocular disparities that can exceed the human limits for fusion.

To summarise the process of identifying the space of binocular correspondences with known surface shape: for a fixed viewing geometry, the entire surface of an object is naturally divided into patches for which stereo matches are smooth. These patches are separated by areas for which no

local solution exists, and while global stereo-matches for the margins $S_L - S_L'$ and $S_R - S_R'$ may exist, they should be filtered out.

Using these ideas, the stereo-matches for a specular surface of known geometry and for given viewing geometry can be found computationally using an iterative approach. In particular, we could construct regions of unique solutions S_L' and S_R' and then compute one-to-one correspondence between them. In practice it is usually computationally simpler and more flexible to find all the potential matches first (including global ones) and thereafter filter out the inappropriate global matches. This has the advantage that specific criteria can be used to determine which matches are filtered out, for example, to select only those matches that could in principle be measured by the human visual system. To this end, we have implemented a matching algorithm by constructing a grid of points on surface region S_L that is visible to left eye, and then for each point P_L of that grid, we find (through brute force search), corresponding points $P_R \in S_R$ that satisfy $|\omega_L(P_L) - \omega_R(P_R)| < \gamma$, where γ is the numerical precision with which we can measure non-zero values (Fig. 6). Specifically, correspondence is found by searching for such locations that left and right reflected ray vectors become parallel (within double floating point precision). Note that this approach generalizes to arbitrarily fine grid resolutions (again, within numerical limits).

Given the resulting set of multiple candidate matches for P_R , we find a unique match by selecting the one that is closest to P_L (i.e., minimizing disparity). This method gives us all the stereo-matches. However, not all of these stereo-matches will be fusible. In particular, there are limits on disparity gradients for stereopsis (Burt & Julesz, 1980; Tyler, 1975) and limits on the vertical offsets between matched features (Qin, Takamatsu, & Nakashima, 2006; Van Ee & Schor, 2000). In order to visualise those disparities that fall within these limits, in Figure 6 we show the subset of matches that are likely to be fusible by the human visual system.

This analysis shows that reliable stereo-matches form local patches with smooth and continuous mapping surrounded by narrow areas that are likely to be unfusable. Notice also that unfusable regions correspond to regions of low Gaussian curvature (as indicated by the surface colour map), i.e. regions around inflection contours where at least one of the principle surface curvatures changes its sign and thus surface normal vectors reverse.

3.2. Correspondence is not limited to epipolar lines

As can be appreciated from the vector field connecting binocularly corresponding points on the surface of a 3D object in Figure 6, corresponding points can be shifted with respect to each other in any direction depending on topological properties of local surface patch. This makes the specular disparity field different from Lambertian objects for which disparity vectors are strictly limited to epipolar lines. These omnidirectional offsets are a consequence of the fact that the viewing vectors do not intersect in three-dimensional space, thus points of correspondence are not found along epipolar lines. This is important for image-based methods of finding corresponding points because such methods often rely on an epipolar constraint and would therefore fail with specular reflections. We address these issues in detail in Section 4. The presence of potentially large non-epipolar disparities can dramatically influence fusibility in human stereo vision, meaning that even smooth local matches can present difficulties for stereopsis. Therefore large portions of mirror reflective surfaces may be binocularly undetermined because either

stereo matches do not exist or non-epipolar offsets are excessively large for human stereo correspondence.

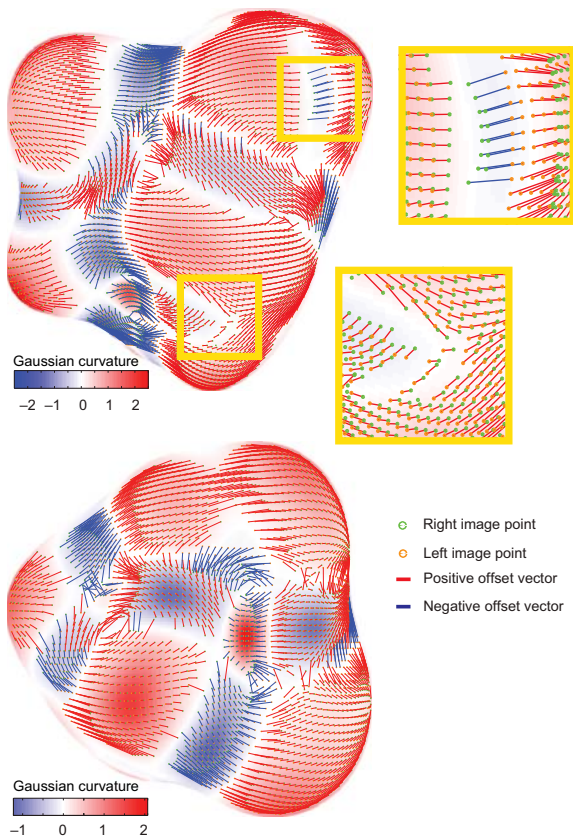


Figure 6: illustration of corresponding points mapped onto an object's surface. We show corresponding points (P_L , P_R) identified by matching reflected ray vectors. A regular grid of points in the left eye image (orange) are matched to points in the right eye image (green). We connect these points to provide a vector flow representation where the colour of the connecting line (red or blue) indicates the sign of the disparity. This vector map is plotted on top of a colour map that shows the intrinsic Gaussian Curvature of the underlying surface. To aid visualization and avoid overcrowding the figure, we down-sampled the matches and displayed only matches with cyclopean separation less than 12 arcmin. The shapes are examples of 'potato' objects (~7cm in diameter), viewed from 50cm with an interocular spacing of 6.5 cm. These were mathematically defined in spherical coordinates, and sample locations are therefore uniform in spherical coordinates (i.e., not regular in the image plane). Sampling in this way allowed us to estimate precise surface normals analytically. This precision was critical because even very small errors in surface normal calculations (which would be unavoidable if we had sampled in the image plane and used numerical methods for surface normal estimation) may lead to large errors in reflected vectors. Our calculations of ground truth matches sampled the visible hemispheres of the shapes (180 x 180 deg) very densely (512 x 512 samples). The results shown here are down-sampled considerably for visualization.

3.3. The relationship between second-order surface structure and disparity magnitude

One final observation to make based on calculating disparities by matching view vectors is that there is an approximate relationship between curvature-like surface properties and the magnitude of the offset between matched locations on the surface, $\xi = P_R - P_L$ (**Fig 6**). This comes about because when there is a high rate of change of surface normal directions, the reflected ray vectors sweep rapidly through the surrounding world, so that a small distance on the surface will encompass a large portion of the environ-

ment. Others have also noted important relationships between second-order surface structure and disparity magnitudes and signs (e.g., (Blake & Brestaff, 1988; Blake & Bülthoff, 1991)).

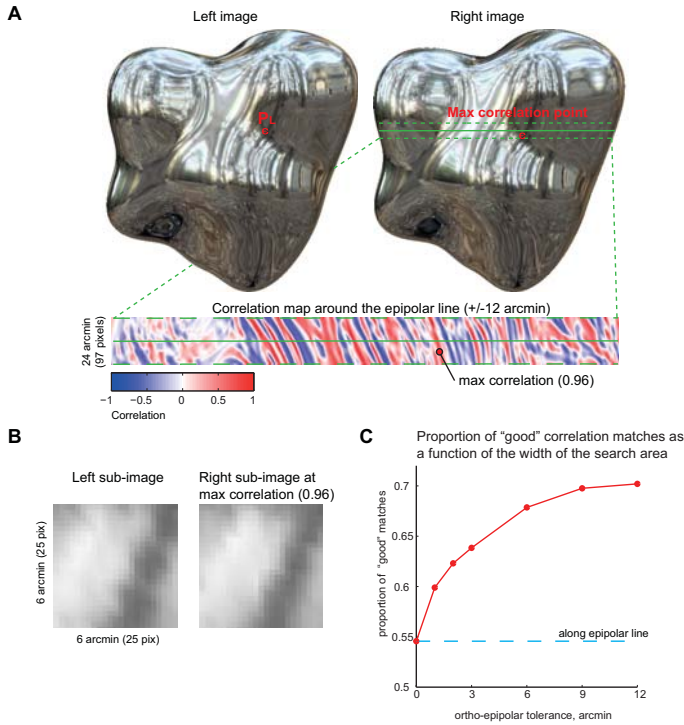
For illustration: consider a fixed point P_L on the surface and suppose we need to find corresponding point P_R . Matching points cannot coincide on the surface, therefore to find a match we need to move P_R away from P_L until corresponding reflected ray vectors match. As we move point P_R away, the viewing vector $\mathbf{v}_R(P_R)$ and surface normal $\mathbf{n}(P_R)$ vary, and, the faster this happens, the lower the distance required before ray vector ω_R meets ω_L . Conversely, for regions of low curvature, the surface normals barely change across local surface patch, meaning that a larger distance over the surface of the object is traversed before a match is found. In the extreme case of a nearly planar surface, disparities become very large indeed. Describing the relationship between surface curvature (an intrinsic object property) and disparity (a viewer property) is necessarily approximate because, as described in Section 2, the view vector depends on both the viewer's location and the surface topology. While this statement is always true when seeking relationships between image cues and physical properties of an object, it is more critical for specular reflections because reflected ray vectors vary faster than the surface normals. Thus approximations relating 3D structure to, for instance, texture patterns are more robust than the equivalent case for specular reflections.

4. A correlation based method for finding stereo-matches in binocular images of a specular object

In the previous section we explained how specular stereo-matches can be found assuming the surface is already known. In general, however, surface geometry is unknown so stereo-matches need to be determined from binocular image pairs. Here we describe an approach for calculating disparities from binocular images of specular objects to provide a proof of computational principle and a cross-validation of the forward model approach we took in Section 3.

To extract disparities, for each pixel in the left image we need to find a pixel in the right image that reflects the same portion of the environment. Consistent with models of standard (non-specular) stereopsis (Banks, Gepshtein, & Landy, 2004; Bolles, Baker, & Hannah, 1993; Cormack, Stevenson, & Schor, 1991; Cumming & DeAngelis, 2001; Filippini & Banks, 2009; Fleet, Wagner, & Heeger, 1996; Harris, McKee, & Smallman, 1997; Kanade & Okutomi, 1994; Ohzawa, DeAngelis, & Freeman, 1990) we take the approach of identifying correspondence based on correlations between image neighbourhoods. Specifically, corresponding pixels are those whose local surroundings correlate most strongly between left and right views. Matching pixel to pixel can be time consuming, and in order to optimize this search conventional algorithms exploit the epipolar constraint, which reduces the area of potential matches to a line (Prazdny, 1983). However, as noted above, specular stereo violates the epipolar constraint and matches can occur anywhere in the image. Given that only local matches make geometrical sense (from a generative perspective, see section 3.1), we suggest searching for solution within $\pm\epsilon$ of the epipolar line, where we have defined $\epsilon = 12$ arcmin for shapes we used.

To find corresponding locations, we used a method where for any left eye image location, P_L , we constructed the corresponding epipolar line in the right image (**Fig. 7a**). We then searched for corresponding points by taking a square sub-image region (length = 6 arcmin / 25 pixels) around sample point P_L in the left image and calculated the pixel-based



correlation for all similar sub-images along the epipolar line in the right image. (The size of the sub-image is somewhat arbitrary; we selected a value that would capture fine detail and wanted to avoid the additional parameters of a multi-scale approach). This created a correlation map centred on the epipolar line (**Fig. 7a**). By default we searched for correspondence by applying a tolerance of ± 12 arcmin (± 48 pixels) around the epipolar line based on experimental results on human fusibility limits for vertical offsets between the two eyes (Qin et al., 2006; Van Ee & Schor, 2000). We selected corresponding points as the peak of the correlation landscape in the right image for point P_L , which typically gave rise to a close match between the sub-images from the left and

Figure 7. Establishing stereo correspondence for specular objects using a correlation method. A) For a given point, P_L , in the left eye image, we search for a match along the corresponding epipolar line in the right eye image. The algorithm correlates graylevel image intensities for a sub-image region in P_L with all the possible locations along the epipolar line in the right eye image. We vary the tolerance around the epipolar line in which to search for corresponding locations. The peak of the correlation map is selected as the corresponding location. B) Identified corresponding locations typically have similar image structure (here, $R = 0.96$). C) Systematically varying the search zone around the epipolar line reveals that allowing some tolerance is important when finding corresponding locations for specular objects. When the search zone is ± 12 arcmin of the epipolar line, the match between the ground truth forward model and the correlation approach approaches the best achievable (i.e., saturating function) for the shapes and viewing situations we have considered. ‘Good’ matches are those that correlation solutions that lie within ± 0.25 arcmin of solution based on matching reflected ray vectors. The function saturations around 0.7 as some portions of the shape are unfusable based on the ray-matching approach – see Fig 6 – so matches are beyond the search zone of 12 arcmin and can therefore never be ‘good’ (i.e., viewers would not be able to extract the disparity information from these locations).

right images (**Fig. 7b**). By systematically manipulating the tolerance value (ϵ) we examined how critical this was in establishing correlation-based matches that were close to (± 0.25 arcmin) those identified based on the ray-matching forward model approach (**Fig. 7c**).

To demonstrate the approach, we rendered stereo-images of specular objects in retinal angular coordinates with high resolution such that 1 pixel = 0.25 arcmin with a 50 cm viewing distance. We converted images into grey-scale by averaging the RGB channels, and then constructed a grid of 104x104 sample points in the left image (**Fig. 8**). Note that the grid is uniform in terms of the angular coordinates of the shape, not in terms of image coordinates, although the precise choice of locations to test is arbitrary: other locations could also have been tested given sufficient mathematical precision (see Fig 6 caption).

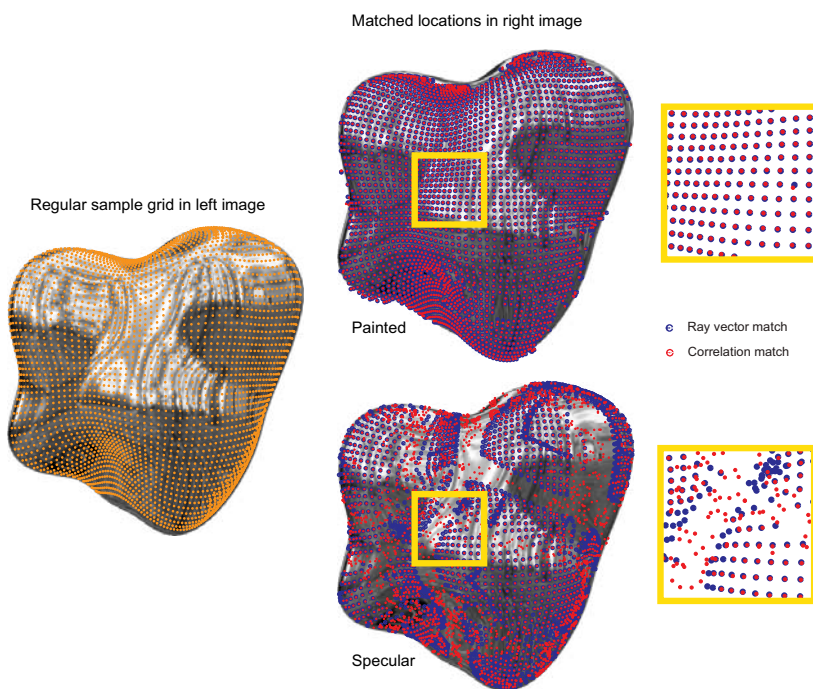


Figure 8. Stereo matches calculated using a correlation-based image method.

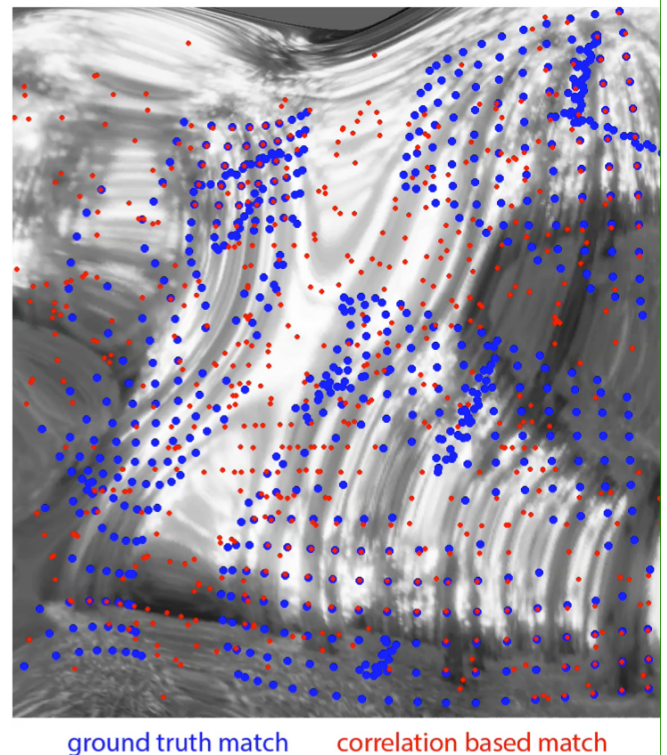
We sought to establish correspondence using an image correlation approach. We start with a regular sample grid in the left eye and then identify corresponding locations in the right eye. (Note that the grid is uniform in the spherical coordinates of the shape, rather than in the image plane, matching our analysis in Figure 6). We show results for a ‘painted’ shape and a specular rendering of the same shape. We superimpose matched locations identified from ray geometry (blue dots) with those identified using the correlation method (red dots). For the painted case, there is very good correspondence between the two. For the specular, there is good correspondence for local surface regions, however in other regions where disparity is undefined (as per the reflected ray analysis with known object geometry) the correlation method produces spurious matches. These regions are likely to pose a similar challenge to the human visual system – see Figure 4. Movie 1 shows the matches for the specular object with different amounts of tolerance for matches with respect to the epipolar line.

As a sanity check on this approach, we first applied this algorithm to a ‘painted’ version of the object (i.e. one with mirror-like monocular appearance, but whose disparities lie on the surface, like a matte textured object). We found very good correspondence with the ground truth depths (**Fig. 8**), indicating that when disparities are well defined, our correlation-based matching algorithm provides good results. We then calculated correlation-based stereo-matches for a specular version of the object, and compared the results with those derived from the reflected-ray based approach (Section 3), which we defined as ground truth for the disparity signals available from these images. The correlation-based approach recovers features of the disparity structure that are similar to the ray-based approach with known geometry. In particular, for well-defined ‘islands’ within the shape there is a good correspondence between the recovered disparities and the ground truth stereo-matches. However, in other portions of the shape there is poor correspondence (i.e. correlation at the best match was low, suggesting residual errors in the match). Importantly, these regions corresponded to the locations where disparity was undefined and for which no local solution exists in the ground truth. Thus, the correlation-based method with unknown geometry, and reflected-ray approach with known geometry yield broadly similar results.

The correlation-based method we have used is not sophisticated and is slow (97 times slower than it would take for a Lambertian object with standard epipolar constraints). Many other existing stereo algorithms are likely to be more efficient or accurate (for a recent review of such work which assesses the relative merits of different algorithms on benchmark tests, see (Baker et al., 2011)). However the results of our method correspond well to the ground truth, which demonstrates that in principle stereo-information from mirrors is available for machine vision, through a simple generalization of standard stereo-matching approaches. We consider this a proof of concept that specular disparities can be robustly calculated from images of unknown surfaces. The most conceptually significant difference between our approach and existing methods—which are optimized for standard stereopsis—is that we do not restrict search for matches to epipolar lines. To evaluate the extent to which this alters performance, we measured how close the image-based matches are to ground truth as a function of the size of the tolerance around the epipolar line (**Figure 7c; Movie 1**). Enabling matches that deviate from the epipolar line is clearly very important for achieving accurate matches, although increasing search window beyond the limits of human fusibility would yield diminishing returns (the curve saturates).

One other advantage of a simple correlation-based matching approach is that additional information can be acquired from the magnitudes of the correlations associated with the best match. Where the correlation is very high, the quality of the match is good, but where the view geometry leads to substantial deviations in the pattern of reflections between the two eyes, the maximum correlation will likely be lower, indicating a poorer quality match. In principle human or machine vision could exploit information about the quality of matches to weight the signals derived from different surface locations. Indeed, we have argued that the visual system prefers to interpolate across regions where the disparities are too unreliable, yielding smoother estimates of the disparity field (Murry et al., 2013). For the remainder of the article we use the ground truth solutions based on known geometry.

Orthoepipolar tolerance 0 arcmin



Movie 1. To complement the results shown in Figures 7 and 8, this movie illustrates how changing the tolerance for non-epipolar matches changes the spatial consistency between matches based on ray geometry and image correlation. The image patch corresponds to the local region shown in Figure 8. Different frames of the movie show difference positive and negative tolerances around the epipolar line – from zero (strict epipolar) to a search zone of 24 arcmin centered on the epipolar line (12 arcmin tolerance). Notice that as the tolerance increases, more red dots line up with blue dots. However, there is never a perfect match because for some locations on the shape disparities become very large or are undefined – see Section 3.

5. Ortho-epipolar distances and their potential use as a shape cue.

For a Lambertian object, binocular correspondence falls along epipolar lines, but as discussed in Section 4, for specular surfaces this does not have to be true. Specular stereo matches generally fall some distance away from the epipolar line, depending on the orientation of the offset vector $\xi = P_R - P_L$ between the corresponding points on the surface. The closer this vector is to coplanar with inter-ocular axis, the closer the stereo-match is to the epipolar line. Intuition suggests that the non-epipolar signals are not randomly distributed across the surface but are systematically related to specific geometrical properties. In this section, we describe how the relationship between surface and view geometry leads to unusual patterns of disparities, quite unlike those seen with standard matte/textured surfaces. A vision system could in principle use these non-epipolar disparities to infer additional information about the shape of the surface that generated the signals.

To start, consider a generic local surface patch, which has different curvatures in different directions (i.e. it is non spherical). Along the direction of highest curvature, corresponding reflected ray vectors tend to match up quickly and thus projection of the offset vector onto the direction of maximum principal curvature is likely to be smaller than its projection onto the direction of minimal curvature (see Section 3.3).

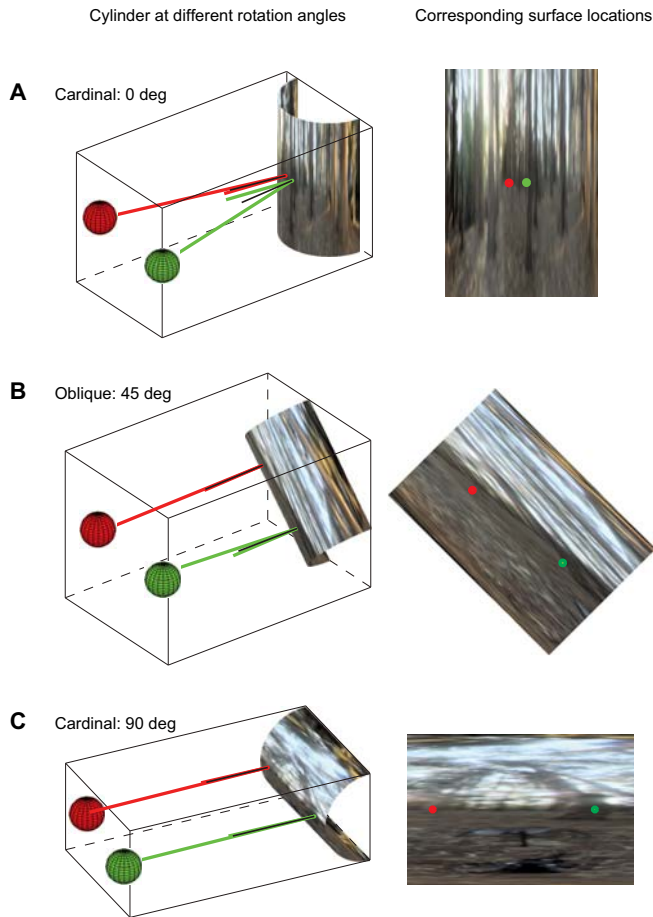


Figure 9. The orientation of disparity vectors for different viewing geometries. Here we show corresponding reflection locations on the surface of a cylinder at different rotation angles. In (a), the direction of minimum curvature is aligned with the interocular axis and the resulting disparity contains no vertical component. At oblique orientations (b) the vertical component of the offset between the two eyes can be considerable. Intuition might suggest that the orientation of the offset vector between corresponding surface locations is related to surface curvature, however part (c) demonstrates that matches are horizontal when the direction of zero curvature is orthogonal to the inter-ocular axis. Therefore a formulation that incorporates viewing geometry is needed to capture the relationship between the magnitude of ortho-epipolar disparity components and the viewed shape.

Thus, at first glance one may think the offset vector ξ should be oriented primarily along the minimal principal curvature (Fig 9a and b). However, as shown in Figure 9c, when the minimum principal axis is vertical (i.e. zero curvature is orthogonal to the inter-ocular axis), the matches are also hori-

zontal, because the depth variations reduce to one dimension. This demonstrates that surface curvature influences deviations from epipolar geometry in a way that is very different from matte-textured surfaces. More generally, however, surface geometry alone cannot fully predict the orientation of the disparity vector because reflected ray vectors depend on viewing vectors as well as on surface normals. Thus, the problem of offset vector orientation cannot be formulated in purely object-centric terms such as curvature, but must also include viewing geometry.

In order to determine the extent to which specular surfaces violate the epipolar constraint, we calculated the ortho-epipolar distance (i.e. image distance between the match point and corresponding epipolar line (Read, Phillipson, & Glennerster, 2009)) for every matched pair of points in an image. Figure 10a shows the results. Superimposed contours represent points where viewing vectors intersect and thus ortho-epipolar distance is zero. These singularities are especially interesting because they relate to the orientations of the offset vectors (i.e., the 2D vector between corresponding points in the two eyes' views, for a given vergence angle) and through these to surface and viewing geometries.

In order to incorporate viewing geometry into the analysis, we constructed a depth map of the object from the cyclopean viewpoint and computed the eigenvectors of the Hessian matrix of this depth map (Fleming, Torralba, & Adelson, 2004; 2009). The Hessian matrix captures the rate of change of surface normal as a function of distance in the image, and therefore incorporates information about both viewing and surface geometries. The eigenvectors of the Hessian matrix are orthogonal to one another in the cyclopean image plane, and represent the directions in which surface normal changes fastest (direction of maximum second derivative) and slowest (direction of minimum second derivative) respectively. For example, for a patch of surface that is locally cylindrical, one eigenvector direction runs in a straight line along the axis of the cylinder, whereas the other runs orthogonally around the circular cross-section of the cylinder. If either one of these two eigenvectors is parallel to the inter-ocular axis, then ortho-epipolar distance tends to zero. In other words, if the cylindrical patch is either horizontal or vertical relative to the eyes, then matches for that location will lie on epipolar lines, just like in standard stereopsis. Figure 10 demonstrates this correspondence—note that areas of zero ortho-epipolar distance correspond to zero orientation. In terms of the mirrored cylinders considered in Figure 9, if the axis of the cylinder is either parallel to—or orthogonal to—the inter-ocular axis, then disparities are purely epipolar. In general all other locations tend to have non-zero ortho-epipolar components, unlike standard stereoscopic matches, which always lie on epipolar lines.

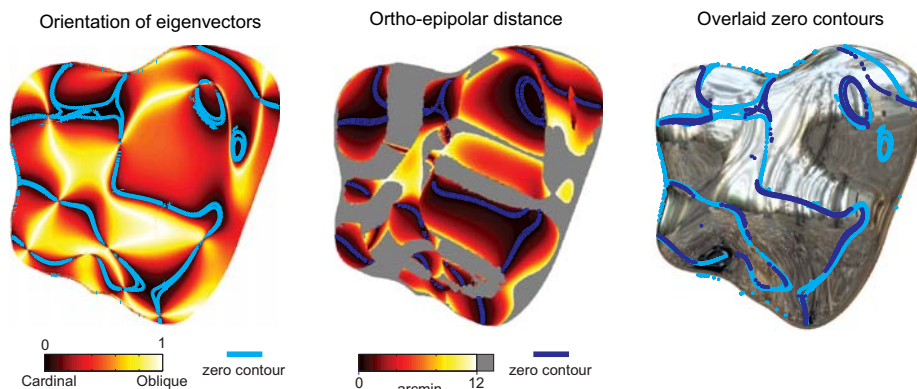


Figure 10. The relationship between surface/viewing geometries and ortho-epipolar distances. (a) A plot showing locations for which ortho-epipolar distances are zero. (b) A plot showing locations at which the orientation of either of the two eigenvectors of the Hessian matrix of a depth map of the object is parallel to the inter-ocular axis (loosely speaking, the directions of the principal curvatures of the surface in a view-centered coordinate system). The correspondence between the two suggests second-order surface properties play a key role in determining where specular stereo conforms to the epipolar constraint.

This observation has one potentially interesting consequence for shape reconstruction from specular disparities. An artificial vision system could in principle use the loci of epipolar matches as additional constraints on the second-order properties of the generating surface at those locations, as it indicates that the eigenvectors must be parallel or orthogonal to the inter-ocular axis. In practice, however, it seems relatively unlikely—but not impossible—that the human visual system makes use of this constraint, at least for the estimation of metric second-order properties. In our previous work we find that where features are easily fused, subjects tend to take the resulting depth estimates at face value, incorrectly interpreting them as the true surface locations (Murty et al., 2013). We suggested that ortho-epipolar components indicate the underlying disparities are unreliable. Within this framework, epipolar matches provide the *most* reliable disparity signals. If the human visual system applied this constraint appropriately, then inconsistencies between depth signals, and the inferred second-order surface constraints should veto, or at least influence the resulting depth estimates, but they do not. This suggests the visual system does not exploit this constraint. Nevertheless, the constraint could prove useful for artificial vision systems.

5.1. Distinguishing specularities from half-occlusions

Unmatchable features or pseudo-matches that infringe epipolar geometry can be created not only by specular reflections, but also by half-occlusions (Da Vinci stereopsis). How then might the visual system distinguish between these two quite different physical causes? Occlusion is substantially more common than specular reflection in the natural environment, so one possibility is that the visual system treats unmatchable features (or non-epipolar pseudo-matches) as evidence of occlusion by default, and that it is only inconsistencies between binocular and monocular cues to occlusion that veto this interpretation. Even in the absence of monocularly visible boundaries, binocular occlusion cues—such as unmatchable (or incorrectly pseudo-matched) line terminators—are strong enough to yield vivid illusory contours, and constrain the orientation and depth of the illusory occlude (Anderson, 1994; Gillam & Grove, 2004; Grove, Brooks, Anderson, & Gillam, 2006; Grove, Byrne, & Gillam, 2005). In general, however, unmatchable features caused by occlusion occur more frequently on the left and right flanks of an object, and tend to be spatially aligned with monocular occlusion cues. In contrast, those created by specular reflections can occur at arbitrary locations in the centre of the object. Indeed, the unmatchable features (or regions with very large ortho-epipolar components) created by specular reflections generally do not lie close to occlusion events, instead occurring in locations where there are clear monocular indications of a continuous surface. Thus, the presence of unmatchable features at locations where monocular cues are inconsistent with occlusion, could provide a reliable indicator that specular reflection (or some other surface-related physical process, like refraction) is the underlying cause.

An alternative possibility is that there is something about the pattern of the unmatchable features themselves that indicates that occlusion is an improbable interpretation. When occlusion is the cause, unmatchable features are typically narrow, elongated areas along the contour, flanked by clearly fusible regions, as the binocularly visible portions of matte surfaces are easily matched, yielding reliable disparity signals. By contrast, with specular reflections, unmatchable regions are not constrained to be elongated in shape, and, more importantly, fusibility usually declines progressively towards the unmatchable region. Unmatchable regions in the

middle of specular surfaces are typically surrounded by areas of partial fusion, with increasingly large disparity gradients or large ortho-epipolar components to the disparities. Thus, the visual system could use both monocular cues and the spatial context of the unmatchable features to determine their origin.

To illustrate these properties, **Figure 11** shows epipolar and ortho-epipolar disparity fields and the epipolar disparity gradient (i.e., the gradients of the epipolar disparities along the epipolar lines) for painted (vIP=0), mirrored (vIP=1) and midway between these stimuli (vIP=0.5). Notice that disparities reach extreme values at the edges of reliable patches, which is in line with the formal definition (section 3) that there is no reliable depth beyond these patches. Notice also the magnitude of the ortho-epipolar disparities can be quite large for both vIP=1 and vIP=0.5, while it is everywhere zero for the vIP=0 case. Gradients of epipolar disparity are also large at the edges of the smooth local patches: these are likely to pose a challenge to the mechanisms of binocular fusion (Burt & Julesz, 1980).

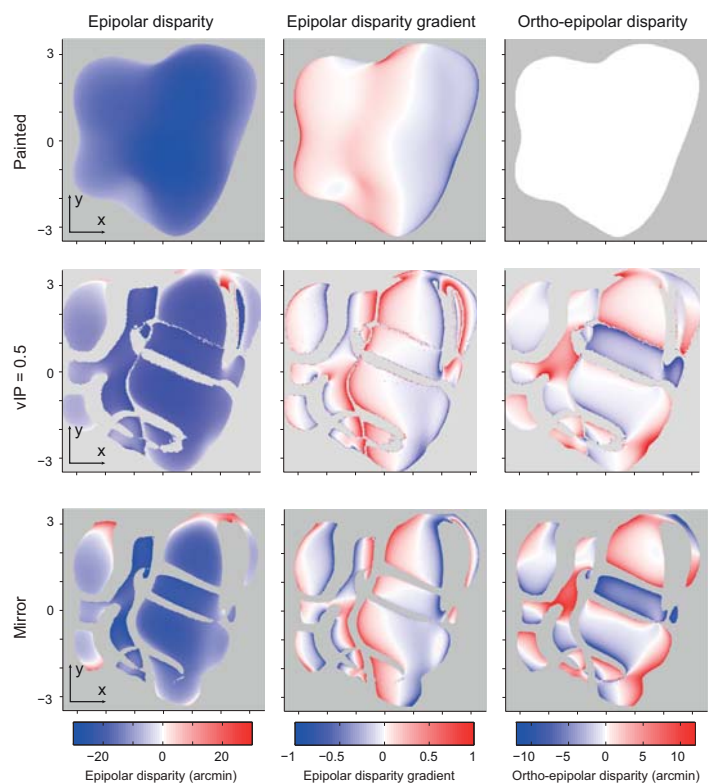


Figure 11. Example disparity fields of an irregular 3D object (a ‘potato’). We show maps of epipolar disparity, disparity gradients along the epipolar lines, and ortho-epipolar disparity for three different vIPs: ‘painted’ vIP=0 (top row), vIP=0.5 (middle row) and ‘specular’ vIP=1 (bottom row) versions of an irregularly shaped object. The object is viewed along the depth axis; x, y image locations are in centimeters. The red-blue color code indicates the magnitude of each quantity (color bars are scaled for each column). Notice that there is a greater range of values for all quantities for non-zero vIP stimuli. This is particularly marked for ortho-epipolar disparity signals. Gaps in the maps are regions for which disparity is undefined or exceeds the fusion limits of the human visual system.

6. Interpretation of stereo-matches as virtual surface depths

6.1. Corresponding ray vectors are typically skewed

Having identified stereo-matches it is useful to explore how depth values could be calculated from the image disparities. For standard stereopsis with matte/textured surfaces and known vergence, calculating depths from corresponding points is straightforward trigonometry. However, for specular surfaces, the image depends on the interaction between the viewpoint and the properties of the local surface patch. This has the important consequence that corresponding vectors are, in general, *skew* and thus do not intersect in 3D space (**Figure 12**). Thus, depth values cannot be trivially derived for a given stereo-match, because there is no unique point of intersection between the two eyes' views. There is, in principle, therefore complete ambiguity about where the depth of the match should lie as correspondence could be established at any point along the view rays for the two eyes (Van Ee & Schor, 2000). Despite this, the human visual system appears able to select from these potentially ambiguous matches giving rise to an impression of binocular depth.

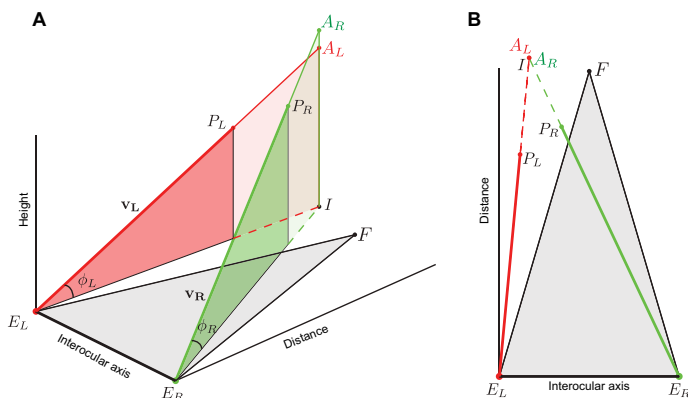


Figure 12. Establishing depth locations for skewed view rays.

Rays reflecting the same portion of the environment are not constrained to epipolar lines meaning that they can pass each other in 3D space without defining a unique point of intersection. Establishing the depth corresponding to the left and right eye views is therefore undefined on the basis of simple trigonometry. We illustrate this in three dimensions. The observer fixates point F . Consider surface locations P_L and P_R that point to the same location in the illumination map. Extending the view vectors (\mathbf{v}_L and \mathbf{v}_R) to determine the depth of the matched feature of the illumination does not define a unique location in 3D space because the vectors pass each other. We therefore establish depth by projecting \mathbf{v}_L and \mathbf{v}_R into the fixation plane where there is an intersection. This intersection location is then projected back onto the view rays to define a virtual point for the left and right eyes (A_L and A_R). The depth of virtual point (A) is defined as the average 3D location of A_L and A_R . Assuming the visual system uses only the horizontal component of the disparity is equivalent to projecting the view vectors into the fixation plane. Figure adapted with permission from the supplementary information of Murry et al (2013).

To arrive at a depth estimate, we have to define a location in space that should be considered the triangulation point. One possible solution is to calculate the *shortest* midpoint between skew viewing vectors (i.e. the point at which they pass closest to one another) and consider such a point as a depth estimate. Although this makes intuitive sense from a geometrical standpoint, it is unclear how a visual system would be able to calculate such a property. We therefore take the approach of projecting viewing vectors into the fixation plane where they must intersect. We can establish matches

based on this projection, and thereafter project the point of intersection back onto the viewing vectors for the two eyes. This establishes two depth locations (one each for left- and right- eye views). We then take the mean location in 3D space as the depth solution. While this strategy may sound convoluted, it is equivalent to estimating depths by ignoring the ortho-epipolar component of the disparity. Measurements of human depth matches for specular objects suggests that this strategy provides a close approximation to human depth perception (Murry et al., 2013). As ortho-epipolar disparity does not indicate depths in standard stereopsis, it is perhaps unsurprising that the visual system ignores it for specular reflections too.

6.2. Characteristics of the virtual surface

Having established points of correspondence between the (generally ambiguous) disparity values provided by specular reflections, we can trace out a virtual surface in depth. In this section, we characterize some of the properties of the virtual surfaces produced by specular objects and report three interesting properties of the depth surfaces that result when viewing specular objects. To illustrate these points we show examples in **Figure 13** of the virtual surfaces for two different types of 3D object: (i) a 'muffin' object that is based on a sphere that has been very subtly distorted by applying low amplitude sinusoidal deviations to the depth profile with the result that the object is convex but has slight 'corners'; (ii) a 'potato' object that is globally convex, but whose surface contains local concavities. For the muffin object, stereo-matches are smooth and thus the virtual surface exists everywhere except at the object edges, i.e. there are no discontinuities in the middle of the disparity field. This is not true for the potato object whose virtual surface contains a number of discontinuities.

First, notice that the properties of the virtual surface can be qualitatively different from the physical surface that generated them. In the case of the muffin object, the physical surface is very close to a sphere and has no concavities, yet the virtual surface contains concavities and a much more pronounced 'rippled' depth structure than we might intuit from looking at the structure of the object. While this might appear surprising at first glance, recall that the virtual surface is a product of reflections that vary twice as fast as the surface normals (Section 2.1). Second, the virtual surface can be highly sensitive to small variations of viewing and surface geometries, especially those parts of the virtual surface that correspond to regions of low physical curvature (because the offset vector is longer there and shift of corresponding points results in larger jumps in depth) while regions that correspond to high physical curvature remain more stable.

A third interesting property of the virtual surface is its piece-wise smoothness, which comes from piece-wise smoothness of the stereo-matches ('Potato' in **Figure 13**). Notice that for convex regions, the virtual surface is typically behind the physical surface in depth, while for concave physical patches it appears in front of the surface. This relationship between the physical surface shape and depths is strict in 2D (where viewing vectors must intersect) but can, under specific (rare) conditions, be infringed in 3D. With increasing curvature of the physical surface, the virtual surface approaches the true surface depths. By contrast, as curvature approaches zero, the virtual depths deviate further and further from the depths of the physical surface: further behind in the case of convex physical surface patches, and further in front for concave patches. This has the important consequence that near inflection points (see footnote 2) of the physical surface, the virtual surface contains a singularity and

undergoes a dramatic jump in depth from far in front to far behind the surface, somewhat like a tangent (tan) function (**Figure 13**). This means that undulating low-curvature mirrors yield extreme depth signals, often outside the range that can be computed by human vision.

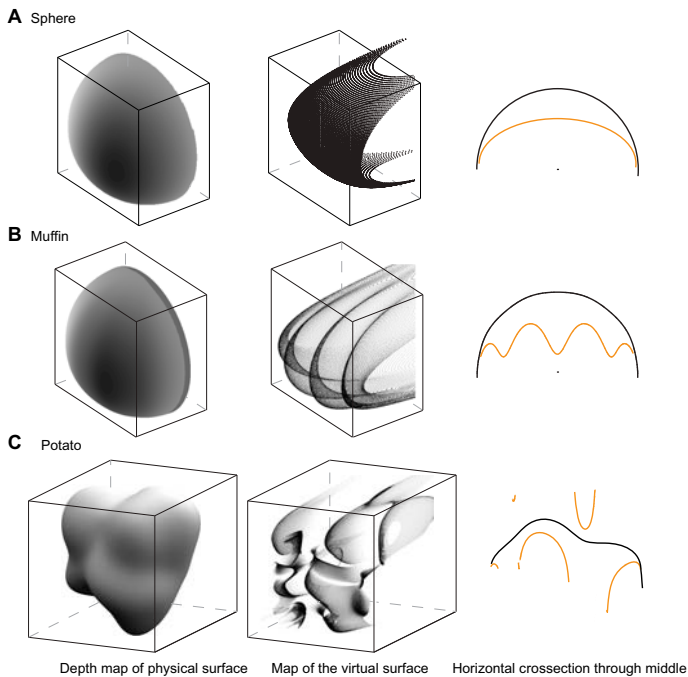


Figure 13. Illustrations of the virtual surface of a sphere, a near-spherical ('muffin') and irregular ('potato') object under different degrees of rotation with respect to the viewer. The depth of the virtual surface defined by specular reflections (central column) becomes increasingly complex as local deviations in the surface are introduced. Even for the muffin which is only a very slight deviation from a sphere (inspect cross section on right of figure) the variation in the depth profile of the virtual surface (orange line) becomes quite pronounced. Viewing distance was 50 cm, and interocular separation 6.5 cm. The sphere had a diameter of 7 cm. The 'muffin' object was created by adding nine sinusoidal bumps ('corners') to a 7 cm diameter sphere in the azimuthal direction, while applying a weighting term to the bumps to ensure that the shape remained convex.

7. Discussion

Our goal with this investigation was to provide a detailed and formal description of specular stereopsis to identify how it deviates from standard stereopsis. In so doing, we also sought to shed light on when and why human binocular surface depth estimation sometimes fails when viewing purely specular surfaces. Our analysis forms a basis for potential future studies both in human and in machine stereo vision. Specifically:

- 1) We introduced a rendering technique based on virtual illumination mapping, which makes it possible to manipulate stereo cues while keeping monocular cues practically unchanged. The method enables the experimenter to continuously interpolate the disparity field between 'mirrored' and 'painted' versions of the stimulus (and indeed beyond) allowing precise control of the conflict between monocular and binocular cues. Here, we used this technique to analyse the structure of specular disparity fields.
- 2) We formulated the solution of the stereo correspondence with known geometry based on matching reflected ray vectors, to establish the ground truth disparities created by specular surfaces. This analysis generalizes previous

work that only considered the behaviour of individual light sources.

- 3) We formally described situations where solutions to the correspondence problem for specular stereopsis do not exist and when there can be multiple matches. This provides crucial insights into when and where human stereopsis should fail when viewing images of mirrored surfaces, expressed in geometrical terms.
- 4) We demonstrated that it is possible to find matching points between stereo-images of a mirrored object using a simple correlation-based method, as long as the matches are not constrained to lie on the epipolar lines, but rather within a 12 arcmin region flanking those lines (consistent with human visual limits fusion for vertical offsets). This demonstrates that simple image-based matching yields disparity fields similar to those predicted from the ground truth structure of the virtual image created by the reflective surface. It also models a simple mechanism through which the human visual system could access both epipolar and ortho-epipolar components of the matches, which we argue are treated as providing different information.
- 5) We showed a relation between specular disparities and surface topology/viewing geometry. Thus, specular disparities can be used to interpret some second-order properties of the real physical surface. Although we argue that the human visual system does not exploit these relationships, they could be used in artificial vision systems.

Previous work has suggested that the human visual system may have internalized the physics of specular reflection (Blake & Bülthoff, 1990; 1991), tacitly implying that it might be able to reconstruct depth from specular surfaces and use the precise locations of reflections to determine surface material properties. This may be true in some qualitative sense—for example, when features lie on the surface in depth they are seen as matte surface markings rather than as highlights. However, we suggest that rather than 'knowing the physics' of specular reflection, many of the limits and problems observers encounter when viewing purely specular surfaces may in fact result from the nature of the disparity signals themselves. In particular, we suggest that the visual system treats the components of the disparity vectors that lie along the epipolar lines as indicators of depth (much as in standard stereopsis), while the orthogonal components may be treated as an intrinsic indicator of the reliability of the depth estimate signal. This approach means that where image regions are unmatchable, no depth estimate results, whereas in locations where matches are epipolar, the visual system treats the depth signals at face value, leading to depth estimates that correspond to the virtual image (i.e., the reflections), rather than the true physical surface itself. In between these two extremes, where features are still fusible, but contain substantial ortho-epipolar components, the visual system may treat the depth estimates as an untrustworthy 'best guess'. Future studies with surfaces that have both reflections and texture, should investigate how the ortho-epipolar components modulate the combination of depth estimates between the accurate and reliable signals from the surface texture with the inaccurate and unreliable signals from the reflections. Our analysis predicts that the depths seen should vary as a function of the specific geometry of the surface and view positions, because these determine the extent of the ortho-epipolar components, and therefore the weight that should be attributed to the depth estimates from the specular reflections.

8. Conclusions

In this paper we have described the process by which images of specularly reflective objects are produced in order to highlight the ways in which specular stereo, differs from the more widely considered matte/textured case. This treatment allows us to make some observations with relevance to artificial matching systems, as well as identify the challenges such images pose to the human visual system. To summarize, the key characteristics of specular stereo we identify are:

- (1) A given feature in one eye may have zero, one or multiple potential matches in the other eye, depending on the surface and viewing geometry. Da Vinci-like unmatchable features routinely occur not just at occlusions, but also at points of inflection on the surface. Surface concavity yields multiple global matches, although constraints on the size of disparities and their gradients can be used to rule out many of these.
- (2) Matches can deviate substantially from the epipolar line, leading to large ortho-epipolar components to the disparity signals. To find matches it is typically necessary to broaden search to a region surrounding the epipolar line. These ortho-epipolar components tend towards zero when the eigenvectors of the Hessian matrix of surface depths (roughly speaking, the principle curvature directions) are parallel to or orthogonal to the inter-ocular axis.
- (3) Corresponding points generally do not yield intersecting rays, so even when correspondence is found, deriving depth estimates from the matches is non-trivial. We suggested the visual system may treat the epipolar component of the disparity signal as a depth estimate, and the ortho-epipolar component as an indicator of the intrinsic reliability of the depth estimate.
- (4) Based on these assumptions, the depth values inferred from specular disparity fields trace out virtual surfaces that fall some distance away from the surface in depth. These virtual surfaces can have qualitatively different structure from the surface that generated them (e.g., convex physical surfaces can yield virtual surfaces with concavities). The virtual surfaces are highly sensitive to view and surface geometry. Smooth physical surfaces can yield virtual surfaces that are discontinuous (piece-wise smooth),
- (5) The depth relationships between the physical surface and its virtual surface are strongly influenced by the physical surface's second-order properties. Depth behaves qualitatively like a tangent function of surface curvature, undergoing a sudden jump—from very far in front to very far behind the surface—as surface curvature transitions from concave via planar to convex. This causes large virtual depth discontinuities around surface inflections.

Together, these properties make specular surfaces highly challenging for vision systems. Our experimental work on human perception of shape and material properties from binocular cues suggests that the visual system has not internalized the specific quantitative relationships between specular reflections and the physical surface that generated them. However, the substantial and systematic deviations from typical behaviour means that specular reflections should often

be relatively easy to identify and exclude where the goal is to estimate true surface depths from stereo signals. In these conditions, interpolation processes are likely to play a key role.

9. Appendix 1

For the analysis in the paper we made the simplifying assumption that illumination is infinitely far from the object. In reality, however, the distance of a point in the environment to the surface of the specular object is finite and therefore the locations mapped out by specular reflections depend on this distance. This is most obvious for flat mirrors where the depth of an object behind a mirror is as far as the real object is in front. Thus, it is important to evaluate the effects of distance on the calculation of the virtual surface.

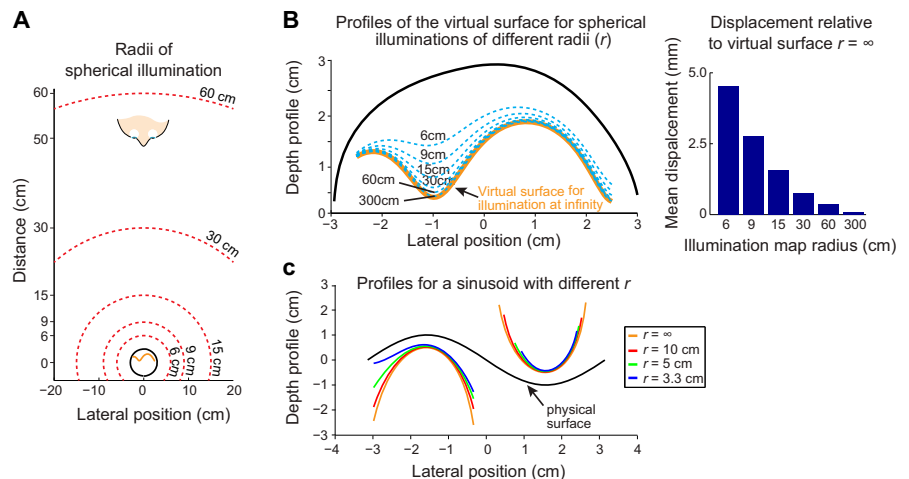


Figure 14. Exploring the ‘illumination at infinity’ assumption. Our analysis in the paper is based on assuming that the environment illuminating the considered objects is infinitely far away. Here we consider different radii of illumination to test the reasonableness of that assumption. (A) Schematic of the viewing geometry and illumination distances. (B) Virtual surfaces produced by illumination spheres of different radii. The solid orange line shows the profile obtained for illumination at infinity. The virtual surface changes systematically as the illumination gets closer to the object (dotted blue lines). We quantify the difference between the mean depth of the virtual surface at infinity and the other illumination distances (bar graph on Right). Beyond a distance of 30 cm, the difference becomes very small. (C) Virtual surfaces for a sinusoidal surface that contains both convexities and concavities. In this case, the difference from illumination at infinity is even smaller than in B. Figure adapted with permission from Murry et al (2013).

To test the importance of the assumption of illumination at infinity and its compatibility with the main conclusions of our paper, we conducted an analysis of the effects of illumination distance on calculated depths. We did this by calculating depth locations of reflections for a simple near-spherical reflective object while using spherical illumination maps of different, finite radii. **Figure 14** shows variation of virtual depth as we change distance to the illumination. Notice that the offset in depth with respect to changing illumination distance is more pronounced for regions of low curvature, while highly curved patches are hardly affected by it. The orange solid line (“baseline”) shows the virtual depth profile for illumination at infinity; the blue dotted lines indicate virtual profiles for illuminations of different radii, and the bar chart shows mean displacement of the virtual surface from the baseline values calculated for infinitely far illuminations. It is apparent from **Fig. 14** that the exact depth of the virtual surface depends on the distance from the reflected environment to the surface of the object. However, if the environment is further than 0.5 m, this difference is negligibly small.

10. Acknowledgments

We thank Andrew Blake for discussions. The work was funded by the Wellcome Trust (grants 08459/Z/07/Z & 095183/Z/10/Z) and the EU Marie Curie Initial Training Network "PRISM" (FP7-PEOPLE-2012-ITN, Agreement: 316746).

11. References

- Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature*, 367(6461), 365–368.
- Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J., & Szeliski, R. (2011). A database and evaluation methodology for optical flow. *Int J Computer Vision*, 92(1), 1–31.
- Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, 24(9), 2077–2089.
- Blake, A., & Brelstaff, G. (1988). Geometry from specularities (pp. 394–403). Presented at the *Proc ICCV*, Tampa, FL.
- Blake, A., & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature*, 343, 165–168.
- Blake, A., & Bülthoff, H. (1991). Shape from specularities: computation and psychophysics. *Phil Trans Royal Soc B, Biological Sciences*, 331(1260), 237–252.
- Bolles, R. C., Baker, H. H., & Hannah, M. J. (1993). The JISCT stereo evaluation. *DARPA Image Understanding Workshop*.
- Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, 208, 615–617.
- Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research*, 31(12), 2195–2207.
- Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Rev of Neuroscience*, 24, 203–238.
- Dąbala, Ł., Kellnhofer, P., Ritschel, T., Didyk, P., Templin, K., Myszkowski, K., et al. (2014). Manipulating refractive and reflective binocular disparity. *Computer Graphics Forum*, 33(2), 53–62.
- Debevec, P. (2008). Rendering synthetic objects into real scenes (p. 1). Presented at the *ACM SIGGRAPH 2008 classes*, New York, New York, USA: ACM Press.
- Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., & Kersten, D. (2011). Visual motion and the perception of surface material. *Current Biology*: doi: 10.1016/j.cub.2011.10.036
- Filippini, H. R., & Banks, M. S. (2009). Limits of stereopsis explained by local cross-correlation. *J Vision*, 9, 8.1–18.
- Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research*, 36(12), 1839–1857.
- Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, 4(9), 798–820.
- Fleming, R. W., Torralba, A., & Adelson, E. H. (2009). *Shape from sheen* (No. MIT-CSAIL-TR-2009-051). Cambridge, MA: Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory.
- Gillam, B., & Grove, P. M. (2004). Slant or occlusion: global factors resolve stereoscopic ambiguity in sets of horizontal lines. *Vision Research*, 44(20), 2359–2366.
- Grove, P. M., Brooks, K. R., Anderson, B. L., & Gillam, B. J. (2006). Monocular transparency and unpaired stereopsis. *Vision Research*, 46(18), 3042–3053.
- Grove, P. M., Byrne, J. M., & Gillam, B. J. (2005). How configurations of binocular disparity determine whether stereoscopic slant or stereoscopic occlusion is seen. *Perception*, 34(9), 1083.
- Harris, J. M., McKee, S. P., & Smallman, H. S. (1997). Fine-scale processing in human binocular stereopsis. *J Optical Soc America: A*, 14(8), 1673–1683.
- Kanade, T., & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9), 920–932. doi:10.1109/34.310690
- Koenderink, J. J., & van Doorn, A. J. (1980). Photometric Invariants Related to Solid Shape. *Optica Acta: International Journal of Optics*, 27(7), 981–996. doi: 10.1080/713820338
- Longuet-Higgins, M. S. (1960). Reflection and refraction at a random moving surface. I. Pattern and paths of specular points. *Journal of the Optical Society of America*.
- Mury, A. A., Welchman, A. E., Blake, A., & Fleming, R. W. (2013). Specular reflections and the estimation of shape from binocular disparity. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1212417110
- Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research*, 30(11), 1811–1825.
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 249(4972), 1037–1041.
- Oren, M., & Nayar, S. (1997). A theory of specular surface geometry. *Int J Computer Vision*, 24(2), 105–124.
- Prazdny, K. (1983). Stereoscopic matching, eye position, and absolute depth. *Perception*, 12(2), 151–160.
- Qin, D., Takamatsu, M., & Nakashima, Y. (2006). Disparity Limit for Binocular Fusion in Fovea. *Optical Review*, 13(1), 34–38. doi:10.1007/s10043-006-0034-5
- Read, J. C. A., Phillipson, G. P., & Glennerster, A. (2009). Latitude and longitude vertical disparities. *Journal of Vision*, 9(13), 11.1–37. doi:10.1167/9.13.11
- Sankaranarayanan, A. C., Veeraraghavan, A., Tuzel, O., & Agrawal, A. (2010). Specular surface reconstruction from sparse reflection correspondences (pp. 1245–1252). Presented at the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. doi: 10.1109/CVPR.2010.5539826
- Templin, K., Didyk, P., Ritschel, T., Myszkowski, K., & Seidel, H.-P. (2012). Highlight microdisparity for improved gloss depiction. *ACM Transactions on Graphics*, 31(4), 1–5.
- Tyler, C. W. (1975). Spatial Organization of Binocular Disparity Sensitivity. *Vision Research*, 15(5), 583–590.
- Van Ee, R., & Schor, C. M. (2000). Unconstrained stereoscopic matching of lines. *Vision Research*, 40(2), 151–162.
- Vasilyev, Y., Adato, Y., Zickler, T., & Ben-Shahar, O. (2008). Dense specular shape from multiple specular flows (pp. 1–8). Presented at the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. doi: 10.1109/CVPR.2008.4587685
- Vasilyev, Y., Zickler, T., Gortler, S., & Ben-Shahar, O. (2011). Shape from specular flow: Is one flow enough? (pp. 2561–2568). Presented at the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. doi:10.1109/CVPR.2011.5995662
- Wendt G, Faul F, Mausfeld R. (2008) Highlight disparity contributes to the authenticity and strength of perceived glossiness. *J Vis*. 22;8(1):14.1–10. doi: 10.1167/8.1.14
- Zisserman, A., Giblin, P., & Blake, A. (1989). The information available to a moving observer from specularities. *Image and Vision Computing*, 7(1), 38–42.